LIBRARY
FISH & WILDLIFE SERVICE
PACIFIC OCEANIC FISHERY
INVESTIGATIONS
OCT 29 1953

NATURE OF VARIABILITY IN TRAWL CATCHES

BY CLYDE C. TAYLOR

FISHERY BULLETIN 83

UNITED STATES DEPARTMENT OF THE INTERIOR, Douglas McKay, *Secretary*
FISH AND WILDLIFE SERVICE, John L. Farley, *Director*

ABSTRACT

The problem of variability in sampling from fish populations by trawl net is examined. The data used in this study were collected on Georges Bank during the summers of 1948–51 by the research vessel *Albatross III*. It is shown that the hypothesis of a random distribution of individuals is not tenable, so that variability from sources other than the sampling technique may be expected. It is shown that, if a type of heterogeneity is assumed such that the mean fluctuates from sample to sample as a Eulerian variable, a negative binomial distribution will result. This distribution applies to the numbers per tow of several species sufficiently well to retain it as a working hypothesis.

It is further shown that the distribution of species and the numbers of individuals represented by each conforms, with good probability, to the logarithmic series distribution. Samples drawn from a logarithmic series distribution are expected to have their numbers distributed in negative-binomial form, so that the latter distribution observed in the data is consistent with theory. On the other hand, the number of species observed in samples drawn from a logarithmic series distribution is expected to be of the Poisson form. This is demonstrated for the 1949, 1950, and 1951 data, but does not apply to the 1948 data, in which an excess of tows with less than 4 species is observed.

Several mathematical models postulating heterogeneity in distribution are discussed. The theoretical interrelation of the Poisson, logarithmic series, and negative binomial distributions is pointed out. The occurrence of all three in the trawl data affords a self-consistent theory of distribution.

From the nature of the relation of the variance to the mean in a negative binomial distribution, it is shown that the amount of information concerning an observed mean can be substantially increased by any procedure which tends to reduce the size of the observed mean. It is recommended that in future sampling the observed mean be decreased both by reducing the length of the tow and by using a trawl net with a smaller opening.

In appendix A, a model for fish populations is postulated which would give rise to the observed distributions. In appendixes B and C, the problem of estimating the parameters of the negative-binomial and logarithmic series distributions are discussed. In appendix D, data on plaice eggs collected by a plankton net are shown to be satisfactorily fitted by a negative binomial distribution.

UNITED STATES DEPARTMENT OF THE INTERIOR, Douglas McKay, *Secretary*
FISH AND WILDLIFE SERVICE, John L. Farley, *Director*

NATURE OF VARIABILITY IN TRAWL CATCHES

By Clyde C. Taylor

FISHERY BULLETIN 83
From Fishery Bulletin of the Fish and Wildlife Service
VOLUME 54

UNITED STATES GOVERNMENT PRINTING OFFICE • WASHINGTON : 1953

For sale by the Superintendent of Documents, U. S. Government Printing Office
Washington 25, D. C. - Price 20 cents

CONTENTS

	Page
Sampling procedure	145
Two-tow station data	147
Distribution of species and numbers of fish	148
Some applications of the logarithmic series to trawl data	150
Distribution of species per tow	151
Use of catch data in the analysis of variance	152
Discussion	153
The distribution theory	153
Application of distribution theory to census problem	155
Summary	156
Appendix A.—A model for fish population distribution	157
Appendix B.—Characteristics and fitting of the negative binomial distribution	158
Appendix C.—The logarithmic series	158
Appendix D.—A negative binomial distribution fitted to number of plaice eggs caught by a plankton net	159
Appendix E.—Proof, smaller sampling unit more efficient with heterogeneity present	159
References cited	162
Tables	163

NATURE OF VARIABILITY IN TRAWL CATCHES

By CLYDE C. TAYLOR, *Fishery Research Biologist*

fundamental problem in deriving the char-
istics of a population from a series of samples
e determination of the magnitude and sources
ariability in successive samples. Such vari-
ty may arise (1) from the manner in which
population is distributed in space and (2)
variations inherent in the method selected
ampling. This paper considers the variability
ciated with catches of fish by the otter trawl
respect to such sources. Aspects of this
lem have been considered by Thompson
8), Hickling (1933), and Barnes and Bagenal
1). Parallel problems for plankton have
treated by Winsor and Walford (1936),
sor and Clarke (1940), Silliman (1946), and
es (1949a, 1949b, 1951).

insor and Walford considered that "the
tion actually observed in plankton sampling
e result of (a) variations in the distri-
n of the organisms and (b) variations in
mode of sampling. Under (b) are included
factors as variations in the speed of hauling,
ations of the haul from the vertical, com-
ness of washing the net, etc." It is obvious
similar factors under (a) and (b) would affect
ariability of trawl hauls.

nce the chi-square distribution of paired
kton hauls did not satisfy that expected for a
son distribution of the organisms, Winsor and
ord examined the variability arising from
rs under (b) and concluded that "a limit to
precision of estimate is set by the relative
bility of the volume sampled."

liman (1946) studied the variability in
kton-tow-net catches of Pacific pilchard eggs.
oncluded that "most of the variability of egg
ts is due to real differences in egg concen-
on."

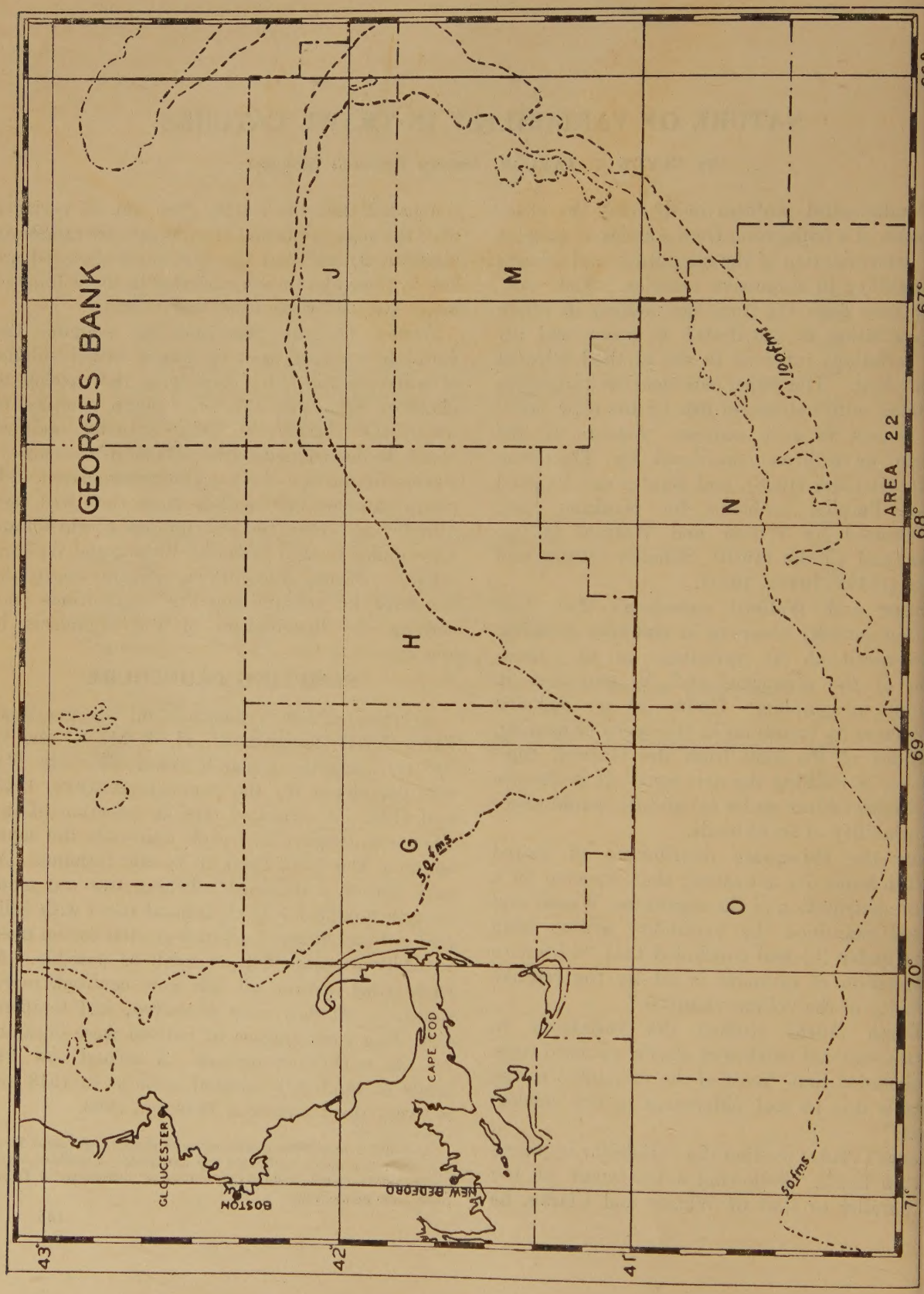
rnes (1949a) studied the variability in paired
kton hauls. Following a treatment of the
parallel to that of Winsor and Clarke, he
concluded that his results "are not at variance with the suggestion that the animals are randomly distributed, and that the variations observed are largely dependent upon variation in the volume of water filtered by the net."

Barnes (1949b) proceeded to examine the variability arising from variations in the volume of water sampled, by comparing the catches of plankton organisms taken by pump samples in which the volumes of water actually sampled could be determined with considerable accuracy. Barnes found that the variability associated with pump samples was no less than that with net samples and that the distribution of chi-square was similar to that found by Winsor and Walford (1936). Barnes' contribution indicates clearly the necessity for reexamining the assumptions concerning the distribution of the organisms in space.[1]

SAMPLING PROCEDURE

A census of fish populations on Georges Bank (Area XXII, subareas G, H, J, M, N, and O, fig. 1), using the research vessel *Albatross III*, was completed for the summers of 1948, 1949, and 1950. A stratified random selection of stations was designed to sample uniformly the depth zones of the bank from 10 to 200 fathoms. At each station a standard half-hour tow was made using a modified No. 1½ Iceland trawl with belly and cod-end liners of 1½-inch-stretch cotton mesh in order to catch fish as small as possible. At each trawl station, all fish were counted, representative samples were measured, and temperature data from surface to bottom were obtained by the bathythermograph. A second tow was made at randomly selected stations in 1948 and at every third station in 1949 and 1950.

[1] Assuming a heterogeneous distribution of plankton organisms, Barnes has shown in a recent study that the distribution of plankton-tow-net catches conforms satisfactorily to the "contagious" distribution of Thomas (1949). See Barnes (1951).


GEORGES BANK
J
M
H
N
O
G
AREA 22
100 fathom
50 fms
50 fms
CAPE COD
GLOUCESTER
BOSTON
NEW BEDFORD
66°
67°
68°
69°
70°
71°
43°
42°
41°

TWO-TOW STATION DATA

nce two tows were made at approximately third of the stations in 1948, 1949, and 1950, series of paired observations furnishes a sure of comparative variability between tows. s Winsor and Walford (1936) show, if the c distribution is Poisson the quantity

$$x^2=\sum\frac{(n-\bar{n})^2}{\bar{n}^2}=\frac{(n_1-n_2)^2}{n_1+n_2}$$

ıld, in a number of random samples, have a n value of 1 and should be distributed in a wn manner.

able 1 shows the distribution of values of chi- are for haddock, whiting, and the common e, and for total fish. The distribution is of same type as those found by Winsor and ford (1936) and Barnes (1949a) for plankton, ı an excess of large values of chi-square.

ı table 2, the totals of chi-square are classified ample size. Although there is some tendency the larger hauls to be associated with high square values, the values are generally high ll levels of catch. While variability in samp- is expected to arise from imperfections asso- ed with the sampling technique, it is clear that ıe data are taken from a population distrib- l at random these imperfections would have e of the grossest kind to account for the ability observed.

s a working hypothesis, the assumption of a lom distribution of plankton organisms does appear unreasonable, even though such an mption seems to have been suspect for some . A similar assumption for the distribution sh, however, appears quite unwarranted, since l catches, fathometer records, and the obser- ons and experience of fishermen indicate that tend to congregate in schools. Trawl catches that these schools are heterogeneous and , while a particular species may predominate, e is ordinarily a variety of species repre- ed. These schools probably differ not only ea but also in density within an area.

table 3, the catches of haddock in tow 1, 2, and tows 1 and 2 combined are tabulated he two-tow stations. It is at once apparent the observed curve of distribution cannot ibly be Poisson unless the mean is less than 1, h it obviously is not.

The observed distribution of frequencies of catch per tow of haddock suggests the negative binomial distribution (Greenwood and Yule 1920). The combined data for the two-tow stations (table 3) were therefore fitted by the negative-binomial distribution.[2] Details concerning the goodness of fit are summarized in table 4.

This development suggested that the basic distribution of the catch-per-tow data might be the negative binomial. It became necessary to examine not only the validity of this hypothesis in the light of the data at hand but also any theoretical reasons why these data should be so distributed.

Negative binomial series were fitted to catches of haddock per standard tow for each of three depth zones: Depth zone 1, 0–30 fathoms; depth zone II, 30–60 fathoms; and depth zone III, more than 60 fathoms. While no particular significance is thought to attach to these groupings, the data were available in these classifications, and these depth zones are presumed to represent a certain homogeneity of conditions. The results of these fittings are presented in table 5. The catches for depth zone III are presented in figure 2.

Similar treatment was accorded the data for the common skate (*Raja erinacea*), the whiting (*Merluccius bilinearis*), and the ocean perch (*Sebastes marinus*). These data are presented in tables 6, 7, and 8. The catches for common skate and ocean perch are presented in figures 3 and 4.

The probabilities associated with the $\Sigma\chi^2$ for these data are summarized in table 9. They are sufficiently high to retain the hypothesis that the basic distribution is the negative binomial. This hypothesis is strengthened, however, if theoretical grounds, conforming to conditions known to apply to the data, may be found.

Let us first consider a Poisson distribution with mean m.[3] Two essential conditions for a Poisson distribution of items are that they be distributed at random and independently of one another. If the items are not distributed independently but tend to cluster in groups,[4] the estimate of m will tend to vary with sampling, and a Poisson distribution will no longer be obtained.

[2] See appendix B.

[3] The development follows in its essentials that demonstrated by Fisher (1941).

[4] Compare the "contagious" distributions of Polya, 1931; Neyman, 1939; Cole, 1946; and Thomas, 1949.

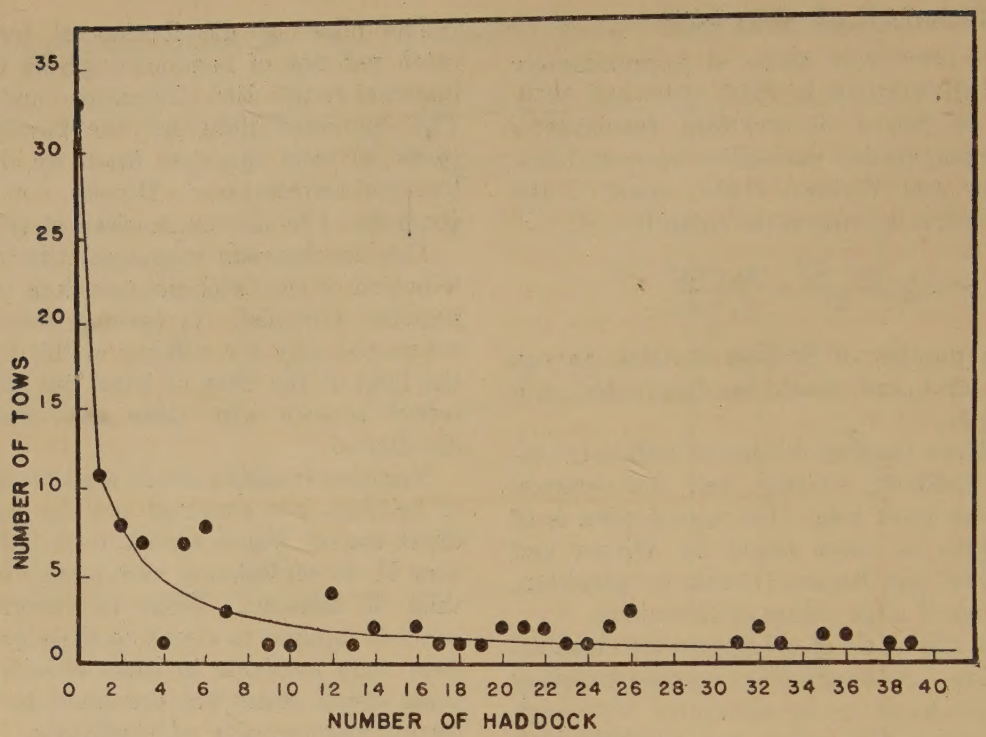


FIGURE 2.—Number of tows, number of haddock, and the fitted binomial distribution.

Fisher (1941) considered the simplest frequency distribution which allows for some variation of m to be the Eulerian distribution with frequency element:

$$df = \frac{1}{(k-1)!} p^{-k} m^{k-1} e^{-m/p} dm \quad \ldots\ldots (1)$$

When m varies in this way, the frequency of occurrence of x units in the sample is:

$$\int_0^\infty \frac{1}{(k-1)!} p^k m^{k-1} e^{-m/p} e^{-m} \frac{m^x}{x!} dm \quad \ldots\ldots (2)$$

leading to the integral of the Eulerian type:

$$\frac{(k+x-1)!}{x!(k-1)!} \cdot \frac{p^x}{(1+p)^{k+x}} \quad \ldots\ldots\ldots (3)$$

which is identical with the standard form of the negative binomial.

In the derivation of the negative binomial distribution, it is important to consider that while the Poisson distribution is that expected for homogeneous material, the Poisson no longer holds if (1) there is variation in the size of the samples or (2) if the material is heterogeneous. The variation (in excess of that expected fo Poisson distribution) observed in plankton trawl sampling has been attributed in the to factors falling under (1), e. g., variations in volume of water sampled. It is clear that v ability may be ascribed equally to heterogen in the population sampled, i. e., "true" samp variation. This heterogeneity, furthermore, not exclude the possibility of a stable [5] mathem cal distribution amenable to statistical treatm

DISTRIBUTION OF SPECIES AND NUMBERS FISH

The number of fish of each species was reco for each tow made over the period 1948–50. logarithmic series having been developed by R Fisher (1943) to account for peculiarities in distribution of species and numbers of individ of each, it appeared desirable to deter whether the species occurring in the trawl sam were so distributed.

[5] By "stable" we mean a distribution whose general characteristics change between successive sets of samples. Heterogeneity, while a cor for a negative binomial distribution, does not necessarily imply the ne binomial distribution as the only possibility.

he trawl-catch data were classified by the e depth zones as were used above. The ıber of species, S, and the number of in- duals, N, were determined for each depth e in each year, and also for each depth zone the 3 years combined, 1948–50. Logarithmic es were fitted to each grouping. Goodness of vas determined in the usual way. The data e grouped so that the number of expected ervations was close to five. After the first terms, the interval of grouping necessary to ain five expected observations increases rapidly hat in general it is not feasible, or necessary, to compute the series beyond the first four or five groupings, these groupings usually including the first 150 to 200 terms.

The results of fitting logarithmic series to the distribution of species and numbers of fish represented by each are presented in tables 12 (a, b, and c) and 13 (a, b, and c), together with the chi-square for each grouping and the probability of obtaining a worse fit by chance for the total chi-square. It is apparent that the hypothesis of logarithmic distribution cannot be rejected.[6]

[6] The relationship of the logarithmic series to the negative binomial is discussed (p. 155 and appendix A). Derivation of the logarithmic series and some of its mathematical properties are presented in appendix C.

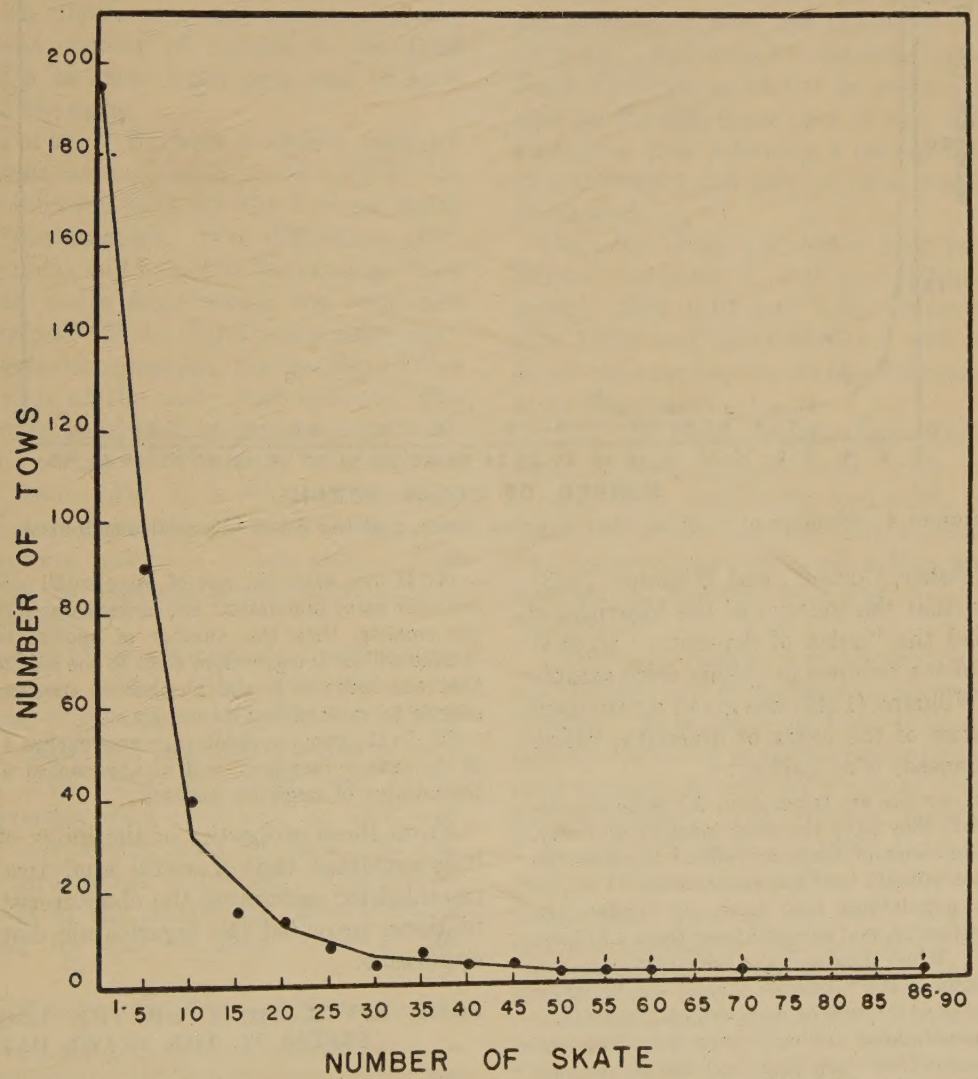


FIGURE 3.—Number of tows, number of common skate, and the fitted binomial distribution.

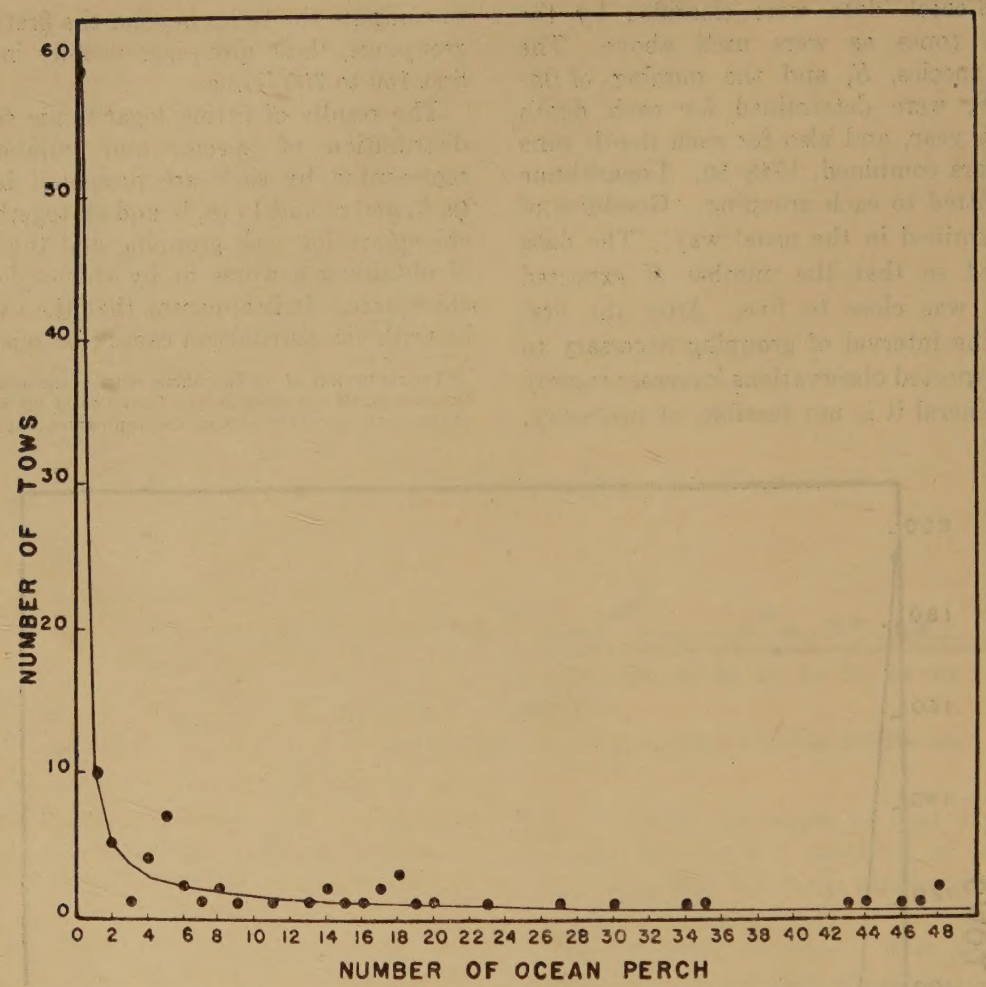


FIGURE 4.—Number of tows, number of ocean perch, and the fitted binomial distribution.

Williams (Fisher, Corbett, and Williams, 1943) has suggested that the index α of the logarithmic series be called the "index of diversity," since it is a measure of the richness in species from sample to sample. Williams (1944) has given a summary of the properties of the index of diversity, which (slightly condensed) is as follows:

(1) If several samples are taken from the same population, not only will they have the same index of diversity, but when two or more of them are added together the combined sample will still have the same index.

(2) Different populations may have, by chance, the same index of diversity, and samples from them will have the same index. When these samples are added together, however, the alpha of the combined sample will be larger than that of either of the two or more original samples.

(3) If two populations are entirely apart, then the index of one sample from each combined will be the sum of the two indices for each sample separately. If there is some overlap, alpha will be larger than either but smaller than the combined total.

(4) If two samples, not of very small size, are tal from the same population, the larger sample being p tir the smaller, then the number of species in the lar sample will be $\alpha \log_e p$ more than in the smaller; it foll that the index α is the number of species added t sample by multiplying its size by e.

(5) In the same population, x approaches 1.0 as the of the sample increases, and n_1 approaches α (where n the number of singleton species).

From these properties of the index of diversi it is apparent that a useful analytical device provided for examining the characteristics of p ulations, provided the logarithmic distributio applicable.

SOME APPLICATIONS OF THE LOGARITHM SERIES TO THE TRAWL DATA

The trawl data collected from 1948 to 1950 Georges Bank were classified by depth zone a bottom type, and the index of diversity and

ance [7] were calculated for each classification. data for depth zones are presented in tables and 15b. Direct comparisons of pairs of es of the indices were made considering dif- nces greater than twice the standard error of r difference to be significant. Since many parisons are possible, one must remember that ut 1 in 20 of such comparisons may be signifi- at the 5-percent level.

one of within-year and between-year compar- s of indices of diversity by depth zones were ificant, so that one may conclude that in hness of species" the depth zones do not differ ificantly either within or between years. This not mean that the species composition does vary with depth zone but rather that the tion of the variety of species to the total ber of fish in each depth zone and in each is about the same.

measure of the differences in species composi- which exist among depth zones may be ob- ed by combining data for the various depth s under consideration. If no differences exist, resulting index will be about the same as those the depth zones from which the combined ple was taken. If the depth zones differ com- ly in species composition, the resulting index be the sum of the individual indices. The h-zone data combined in this way show in- ses in the indices of diversity, but they are er small (table 14). It is evident that there t be many species common to all depth zones fewer species restricted to one alone. This rne out by the observed depth distribution of various species.[8]

ottom samples were attempted at each station re trawls were made. These data were com- d with and supplemented by data from Coast Geodetic charts and classified into four general s: sand, mud, mud and sand, and stony. ces of diversity were calculated for each bot- type in each depth zone for each year and for the three years combined. The indices compared in the usual way, the difference g divided by the standard error of the rence.

nly one difference significant at the 5-percent level was found. Since 30 comparisons were made, we may expect at least one significant difference to appear by chance. These tests do not, then, reveal any evidence of significant differences in the diversity of species either between bottom types within depth zones, or between depth zones within bottom types.[9]

DISTRIBUTION OF SPECIES PER TOW

In the logarithmic series, the distribution of the number of species, S, is Poisson (Anscombe 1950). We may, therefore, further examine the trawl data to find whether the observed distribution of species per tow conforms with theory.

In table 15 the number of species per tow and the number of tows are presented for the years 1948–51. The data for all tows in the Georges Bank area are tabulated in column A. In column B, the first three tows of each cruise and the first three tows following a change from the port to starboard trawl net, or vice versa, have been eliminated.

Data for 1949a and 1950a may be fitted with Poisson distributions with probabilities somewhat greater than 0.10 and 0.80, respectively. The data 1948a and 1951a are fitted with probabilities so low that the hypothesis of a Poisson distribution must be rejected.

Data for 1949b, 1950b, and 1951b are fitted with Poisson distributions with probabilities greater than 0.30, 0.80, and 0.30, respectively. Data for 1948b does not give a satisfactory fit. The observed and expected values are presented in table 16, together with $\Sigma\chi^2$ ($n-2$ degrees of freedom).

There is no reason to reject the hypothesis that the distribution of number of species per tow is Poisson. It is apparent from table 16 that factors are operating in the sampling procedure which tend to result in tows with a small observed number of species. The greatest number of these tows occur in 1948 and 1951; 1948 was the first year of the census sampling. In 1951 a different vessel, the *Delaware*, was used. One is inclined, then, to suspect that factors affecting the efficiency of operation of the trawl net are the cause of a

appendix C for the formula used in calculating the variance.
e statistical significance of the increase in the index through combining be determined by comparing the difference with the standard error difference, since the variates compared are no longer independent.

[9] Preliminary analyses of variance of the transformed catch-per-tow data indicate consistent significant differences in the distribution of particular species by bottom type. The distribution of species by bottom type may be studied by analysis of variance or by χ^2-contingency tables, after a square-root transformation of the number of species per tow.

large number of tows containing a small number of species in these years.

The improvement in agreement of the observations with hypothesis which occurs when the first three tows are eliminated suggests that a dry net does not fish efficiently. This may be true but, on the other hand, the elimination of these tows in the 1950 data makes practically no change at all in the species-per-tow distribution. The trawl data are being studied in an effort to find an adequate explanation of the departure of species-per-tow distribution from theory.

It is evident that if we accept the hypothesis of a Poisson distribution of species per tow, we have a useful method of gauging the efficiency of the sampling procedure. The data presented in tables 15 and 16 suggest that tows are present in the data which ought to be rejected for purposes of further analysis, provided an objective basis for their rejection can be discovered. A Poisson distribution of the occurrence of species in tows also provides means for further study of differences between bottom types, depth zones, or other environmental factors.

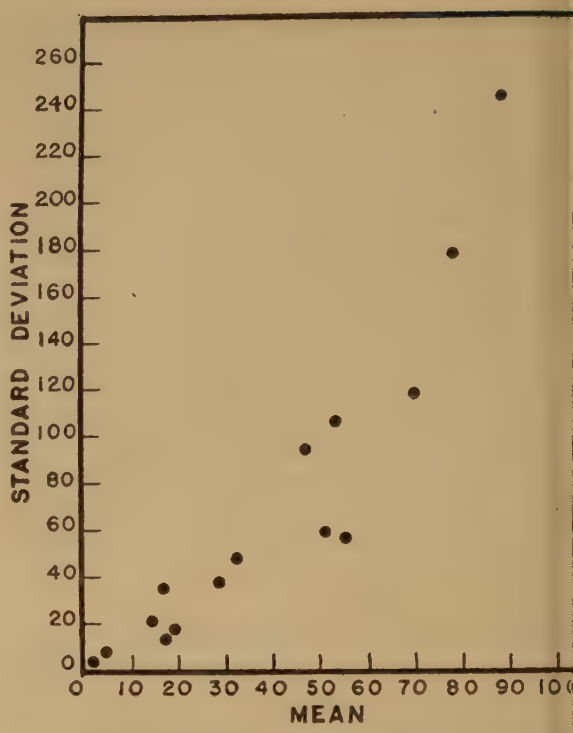


FIGURE 5.—The mean and standard deviation of hadd catches for the subareas and depth zones of Geo Bank.

USE OF CATCH DATA IN THE ANALYSIS OF VARIANCE

When the program for sampling fish populations on the fishing grounds was planned it was anticipated that the technique of analysis of variance could be applied to extract as much information as possible from the body of data to be collected, and the sampling program was designed with this technique in mind. A necessary condition, however, for a valid analysis of variance is that there be a constant residual variance. If the variance changes with the mean level, a suitable transformation is necessary to stabilize the variance.

If the variance can be expressed as a function of the mean m, it has been shown (Curtis 1943) that the appropriate transformation may be determined.

In the early stages of the present analysis, it was found that the standard deviation of the catches per tow was approximately proportional to the mean for many species. Typical examples are presented in tables 10 and 11 and figures 5 and 6 for haddock and whiting catches. The relation between the standard deviation and mean may be expressed

$$\sigma_x = km \quad \ldots\ldots\ldots\ldots$$

where σ_x is the standard deviation of catche with the mean of x equal to m.

It is shown by Bartlett (1947) that for any fu tion $g(x)$, we have approximately

$$\sigma_x^2 = (dg/dm)^2 f(m) \quad \ldots\ldots\ldots\ldots$$

so that if σ_x^2 is to be constant, K^2 say, we n have

$$g(m) = \int \frac{K dm}{\sqrt{f(m)}} \quad \ldots\ldots\ldots\ldots$$

From equations 4 and 6 we have:

$$f(m) = \sigma_x^2 = k^2 m^2$$

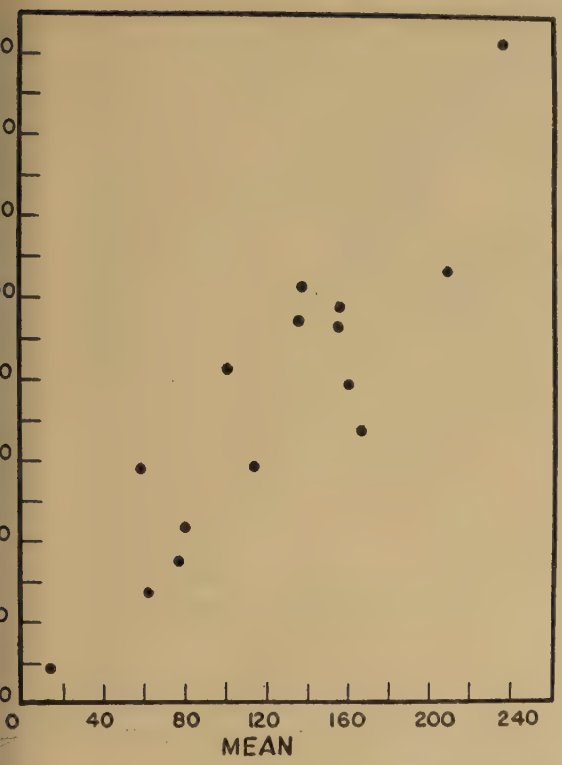
and

$$g(m) = \int \frac{K dm}{\sqrt{k^2 m^2}}$$

or

$$g(m) = (K/k) \log m + C \quad \ldots\ldots\ldots\ldots$$

where C is a constant introduced by integratio



RE 6.—The mean and standard deviation of whiting tches for the subareas and depth zones of Georges nk.

quation 7 indicates a logarithmic transforma- of the catch data. This transformation been frequently employed in the past in the tment of both plankton and trawl data nsor and Clarke 1940; Silliman 1946; Barnes, 0a, 1949b, 1951).

has been demonstrated above, however, that hypothesis of a negative binomial distribution luates the trawl data satisfactorily. The tion between the mean and variance for a tive binomial distribution is

$$\sigma_x^2 = m + m^2/k \quad \ldots\ldots\ldots\ldots (8)$$

lying equation 6 we have

$$g(m) = \int \frac{K dm}{\sqrt{m + m^2/k}}$$

ing to either a $\sinh^{-1}$ or logarithmic integral. combe (1948) shows that for a negative bi- ial distribution with variable x, exponent k, and mean m, the appropriate forms of the transformation are

$$y = \sinh^{-1}\sqrt{\left(\frac{x+C}{k-2C}\right)} \quad \ldots\ldots\ldots (9)$$

and

$$y = \log (x + \tfrac{1}{2}k). \quad \ldots\ldots\ldots (10)$$

If m is large and k greater than 2, the optimum value of C (equation 9) is roughly $\frac{3}{8}$. With k less than 2 but greater than $\frac{3}{4}$, and m large, the simpler transformation (10) may be used.

The negative binomial distributions which have been fitted to the catch-per-tow data have, generally, a fairly large m, but the value of k is, in all instances, less than $\frac{3}{4}$. Under these circumstances, it is apparent that the difference between the transformations indicated by equations 7 and 10 is slight. In applying equation 7, it is customary to use the empirical transformation $\log (1+x)$ to avoid difficulty with zero observations. Equation 10 has the advantage of assigning a numerical value to zero observations, based on properties of the observations themselves and may be preferable in some cases.

DISCUSSION

The distribution theory

It has been shown that the distribution of catch per tow for various species of fish caught by the research vessel *Albatross III* over a 3-year period on Georges Bank conforms to the negative binomial distribution. On the other hand, the distribution of species and numbers of individuals represented by each conforms satisfactorily to Fisher's logarithmic series distribution.

Anscombe (1950) lists seven 2-parameter distributions constructed on models postulating various types of heterogeneous Poisson sampling. Considering only the catch-per-tow data, there appear to be three of these distributions, in addition to the negative binomial, which might be applied: the Neyman Type A, the Polya-Aeppli, and the Thomas, all three being of the so-called contagious type. These distributions are examined, since the trawl catches may represent samples from populations so distributed. The three distributions are similar in supposing a Poisson distribution of colonies or groups, but they differ in the assumptions concerning the distribution of items within groups. In the Neyman

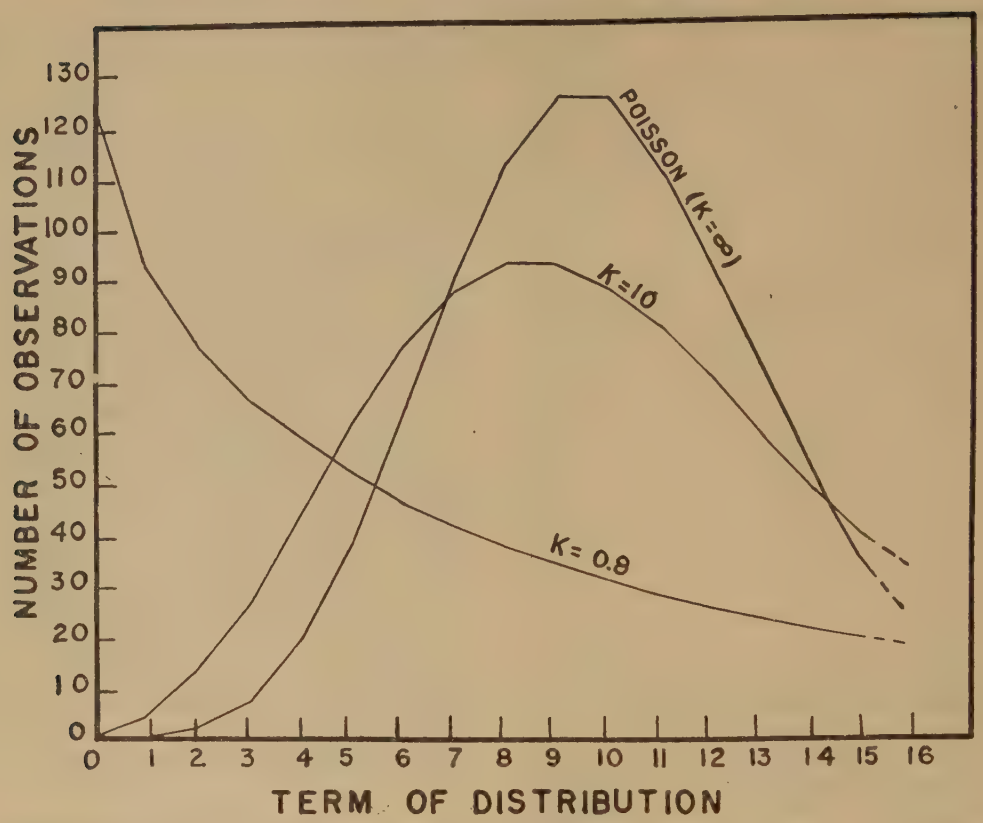


FIGURE 7.—The negative binomial distributions compared to the limiting form, the Poisson. Each distribution ha mean of 10 and consists of 1,000 observations.

Type A distribution, the number of colonies or groups per sample has a Poisson distribution with mean m_1, while the number of individuals per group has a Poisson distribution with mean m_2. The Thomas distribution is similar, the number of individuals per group assumed to be one plus an observation from a Poisson distribution with mean m_2. The Polya-Aeppli distribution, however, supposes that the number of individuals per colony has a geometric distribution; it has, therefore, some interest in connection with sampling from growing populations (Anscombe 1950).

In the absence of a series of equally spaced modes, Anscombe (1950) points out, one may reasonably feel reluctant to use the Neyman or Thomas distributions. Such modes are not evident in the trawl data.

The choice between the negative binomial and the Polya-Aeppli distributions lies less in consideration of the goodness of fit obtained in applying them than in the various theoretical considerations which might justify their use. Kendall (1948) shows that colonies established simultaneously from single ancestors and unaffected by immig tion will, after a fixed lapse of time, have a c tribution of population sizes resulting from grow at geometric rates equivalent to the differe between constant reproduction and mortal rates. This is the model for the Polya-Aeppli d tribution provided that the distribution of t ancestors establishing each colony was rando If, on the other hand, the distribution of ancest does not occur simultaneously throughout t area of observation but is uniform in time, t distribution of colony sizes resulting will be th of the negative binomial, provided that reprod tion and mortality rates are constant. Kenda analysis indicates that the index of the result negative binomial, k, is intrinsically associat with the reproductive power of the species.

These models for the Polya-Aeppli and t negative binomial distributions appear ov simplified in considering the application of th to fish populations. The origin of schools of f can hardly be supposed to occur simultaneou throughout the area of observation or to arise fr

progeny of a single ancestor. The condition niform distribution in time possibly is approxi-ed, so that one may be inclined to favor the ative binomial as the more likely, although an ntial of Kendall's model is the self-duplication andom individuals at random times.

stronger argument in favor of the negative omial distribution is a relation among the ative binomial, the logarithmic series, and the sson distributions which has been pointed out Quenouille (1949). It is demonstrated that in apling from a logarithmic series, the probability ribution of the number of individuals in dom samples is the negative binomial. Que-ille further demonstrated that if any two of three distributions hold, the third distribution mplied (see appendix A).

logarithmic distribution of species and indi-uals is demonstrated to hold with reasonable bability for the trawl data, and it is further wn that the hypothesis of a negative binomial ribution satisfactorily conforms to the dis-ution of numbers per tow, as theoretically uired, given a logarithmic series distribution n which samples are drawn.

The implied Poisson distribution can occur in eral ways, some of which are suggested by the dels considered above. The Poisson distribu-n of the number of species per tow suggests the othesis that the various species are distributed andom and move independently of one another. s hypothesis does not require that the distribu-n of a species be random, since a species occurs y once, no matter what its abundance in num-s, in the sample. The assumption of a Poisson ribution of species appears sufficient to com-e the theory, especially when we have a certain ount of evidence that this distribution actually ds. Alternatively, we might suppose that the urrence of schools or concentrations of fish is dom. This may, in fact, be true but it cannot demonstrated at present because of lack of the essary observations.

lication of the distribution theory to the census problem

The problem of variance.—The primary purpose he census cruises made on Georges Bank was estimate and define the abundance and dis-ution of the fish populations inhabiting it. s not the purpose of this paper to extend the ribution theory to numerical estimates of populations, but it is pertinent to examine certain aspects bearing on them.

In populations with individuals distributed at random and independently of each other, it is well known that the variance is equal to the mean. In a negative binomial distribution, the relation between mean and variance is

$$\sigma_x^2 = m + m^2/k \quad \ldots\ldots\ldots\ldots (10)$$

One sees at once that the departure of the variance of the negative binomial from that of the Poisson depends on the value of k; the variance, as well as the distributions themselves, becoming identical as k tends to infinity. On the other hand, if k is less than one, the variance is never less than the square of the mean and may be of considerably greater magnitude.

The high variance associated with the mean of a negative binomial distribution introduces a most serious problem in estimation. Since k is an intrinsic property of the population being sampled, it is independent of the sampling method and its estimate is subject only to sampling variation. The mean, however, depends on the size of the sampling unit, e. g., in the census sampling, the size of the trawl net and the duration of the tow.

Taking as a measure of the precision of sampling the reciprocal of variance, the amount of information of a mean based on n observations is n/σ_x^2. It is clear that the amount of information can be increased by increasing the number of observations, by decreasing the variance, or both.

Examining the relation between mean and variance (equation 10), we see that any change in sampling method which increases the mean will increase the variance even more rapidly. It has been suggested that the trawl data could be "smoothed out" by using a smaller net and increasing the length of the tow. Increasing the length of the tow would tend to increase the mean, while decreasing the size of the net would tend to decrease the mean. If the two effects were of such magnitude as to cancel each other, the variance would, of course, remain unchanged.

On the other hand, it will be seen that changes in sampling which reduce the mean will also reduce the variance, and for a given number of observations, will increase the amount of information given about the mean. The obvious ways to

reduce the mean are to reduce the length of the tow and to reduce the size of the net within practical limits.

We have, for example, in depth zone II, 171 tows in which the mean catch of haddock was 130. The value of k, calculated from the maximum-liklihood equation, is 0.25 (0.2542). The expected variance is

$$\sigma_x^2=130+\frac{130^2}{0.25}$$

or 67,730. Let us suppose that by decreasing the length of the tow and the size of the net each by one-half, we reduce the observed mean by one-fourth so that it is now 130/4 or 32. The expected variance is now $\sigma_x^2=32+32^2/0.25$ or 4,128. The amount of information has been increased by 67,730/4,128 or about 16 times. Reduction in the size of the net and the length of the tow would doubtless be accompanied by other practical advantages, such as a smaller crew to handle the net, greater ease in handling, more data on the smaller catches, and the execution of a greater number of tows within a given time period.

It is to be pointed out that, in the final analysis, we are interested in the precision of an estimate of the total population in the area sampled. The mean we observe in a series of samples with any given sampling unit is, of course, the basis for this estimate. This problem, with particular reference to a heterogeneous population, is examined in appendix E.

Ecological relations.—Since the numbers of each species occurring in the trawl catches were recorded, we have in these data, in a sense, a third dimension of observation: the relation of the species to each other and to various environmental conditions, not only in qualitative but also in quantitative terms. While some of the properties of the index α of the logarithmic series have been pointed out above, the primary purpose in the present paper has been to examine the reality of the logarithmic-series distribution of species and individuals.

Although the analysis of the indices of diversity by bottom type and depth zone within and between years does not reveal any significant differences in richness of species, an entirely different result is obtained if one uses Fisher's formula for variance (7, appendix C) rather than that given by Anscombe (8, appendix C). In our data Fisher's formula gives a variance roughly one-tenth as great as that given by the formula we have used. As a result many significant differences appear between bottom types, with an apparent tendency for these differences to diminish with depth. With Fisher's formula there still appear no significant differences between depth zones. As Anscombe points out, the conditions under which Fisher's formula is applicable are rather specialized. It is possible that closer examination of our data may reveal that some of the specialized conditions are met.

An interesting relation which may be of practical use after further study of the data is given by the formula

$$S=\alpha \log_e (1+N/\alpha) \quad \ldots\ldots\ldots (1$$

When N is large compared to α, we may write

$$S=\alpha \log_e (N/\alpha) \quad \ldots\ldots\ldots\ldots (1$$

Making the necessary assumption that the density of population within the total area is the same as that within the area actually sampled, we should expect to observe a total of S_t species if we sampled the entire area. Designating as the number of species observed in the sampled area, A_s say, the expected increase in species if we sampled the entire area, A_t, is

$$S_t-S_s=\alpha \log_e (A_t/A_s) \quad \ldots\ldots\ldots (1$$

from which we can readily estimate S_t. Substituting S_t and the sample estimate of α equation 12 we can then estimate N_t, the total number of individuals within the total area of the species. Limits of reliability of estimate may be set from the standard error of α. It is pointed out here merely that the method may be of some use when the area under consideration is not too large as compared with the area sampled, when the sampling is done within a fairly short period, and when the assumption of uniform population density can reasonably be made. Under these conditions Fisher's formula for the variance of S would be applicable.

SUMMARY

Analysis of the census-trawl data collected during three summers on Georges Bank shows that the hypothesis of random distribution

fish of a given species is not tenable. Two oretical distributions, developed from models itulating heterogeneity or aggregation, are wn to apply to the observed data with suf- ently high probability to retain them as work- hypotheses. One of these, the negative omial distribution, describes satisfactorily the tribution per tow of individuals of a species. e other, the logarithmic series distribution, is wn to describe the distribution of individuals d species.

The occurrence of species in successive samples shown to have a Poisson distribution, and this, theory, is sufficient to account for the loga- hmic distribution of species and individuals ob- ved. It is pointed out that various authors ve shown that the occurrence of numbers of a rticular species drawn in samples from a loga- hmic distribution will be of the negative-binomial rm. The observed distributions are therefore f-consistent.

A high variance associated with the mean tch per tow is shown to be a necessary con- quence of the basic heterogeneous distribution the numbers of fish. The agreement of ob- rved variances with theoretical variances is fficiently good to indicate that the variance due variations in the sampling technique is of atively minor importance.

It is shown that the variance may be reduced d the amount of information increased by creasing the size of the observed mean, and it is ggested that this may be accomplished by ducing the length of tow and by reducing the e of the trawl net.

Appropriate transformations for use of the nsus-trawl data in analysis of variance and in her statistical procedures are indicated.

APPENDIX A

nodel for fish population distribution

The following model is an attempt to postulate nple conditions which would lead to the mathe- tical distributions observed in the trawl-sam- ng studies. It is a restatement, in terms of ecies and numbers of fish, of mathematical rela- ns which have been noted in somewhat more neral terms by Fisher (1941, 1943), Kendall 48), Quenouille (1949), and Anscombe (1950). e have borrowed freely from these authors.

Let us postulate that S species are distributed at random within the area of observation, each species moving independently of others, so that the distribution is Poisson with the probability of observing exactly n species in any one sample

$$P(n \text{ species}) = \frac{e^{-m}m^n}{n!} \quad \ldots\ldots\ldots (1)$$

Let us postulate further that the number of individuals represented by each species is not random but rather that the individuals of each species tend to aggregate so that the mean m varies from sample to sample in a Eulerian distribution. The probability of observing a sample of size r is, then,

$$P(r \text{ individuals}) = \frac{(k+r-1)!}{(k-1)!r!} \cdot \frac{p^r}{(1+p)^{k+r}} \quad \ldots (2)$$

which is the standard form of the negative binominal with index k and parameter $p/1+p=x$ (Fisher 1941; Quenouille 1949). Letting k tend to infinity and excluding the first term as unobservable, a logarithmic series distribution is obtained (Fisher 1943) where the probability of observing r individuals of one species is

$$P(r \text{ individuals of one species}) = a\, x^r/r \quad \ldots (3)$$

or the coefficient of t^r in

$$-\alpha \log_e (1-xt)$$

$$\text{where} \quad x = \frac{p}{1+p} \text{ and } \alpha = -\frac{1}{\log_e (1-x)} \quad \ldots\ldots (4)$$

The following relations result (Quenouille 1949):

P (r individuals) $= P$ (n species) $\times$ P(r individuals in n species) $\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$

From equation 4 the probability of observing r individuals in n species is the coefficient of t^r in

$$[-\alpha \log_e (1-xt)]^n$$

Rewriting equation 5 we have

$P(r \text{ individuals})$

$$= \text{coefficient of } t^r \text{ in } \sum_{n=0}^{\infty} \frac{e^{-m}m^n}{n!} \times [-\alpha \log_e (1-xt)]^n$$

$$= \text{coefficient of } t^r \text{ in } \exp [-m - \alpha m \log_e (1-xt)]$$

$$= \text{coefficient of } t^r \text{ in } (1-xt)^{-\alpha m} e^{-m}$$

$$= (1-x)^{\alpha m} \frac{(\alpha m+r-1)!}{(\alpha m-1)!r!} x^r, \text{ since } (1-x)^{-\alpha} = e. \quad (6)$$

Equation 6 is the same as the (r+1)th term in a negative binomial series with parameter x and index α m.

It is thus seen that the two basic assumptions necessary to obtain the observed distributions are (1) the species are distributed at random and move independently of one another, and (2) the individuals within each species tend to aggregate so that the mean m is a Eulerian variable from sample to sample.

APPENDIX B

Characteristics and fitting of the negative binomial distribution

The negative binomial distribution is described by two parameters, the mean m and the index k. The variance of the distribution is

$$m+m^2/k \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

The expected frequency of zeros is

$$P_0=(1+m/k)^{-k} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

and the chance of observing any positive count r is

$$P_r=\frac{k(k-1)\ \ldots\ (k-r+1)p^r}{r!}\ (1-p)^{k-r} \quad \ldots (3)$$

or, more conveniently, letting $p'=-p$, $k'=-k$,

$$P_r=\frac{k'(k'+1)\ \ldots\ (k'+r-1)}{r!}\ p'^r(1+p')^{-k'-r} \quad (4)$$

Haldane (1941) shows that the log-likelihood equation for (4) is

$$N[\log_e (k'+m)-\log_e k']$$

$$=\frac{n_1+n_2+\ \ldots\ +n_R}{k'+1}+\ \ldots\ +\frac{n_R}{k'+R-1} \quad \ldots (5)$$

where n_r is the observed frequency of r, and R is the maximum value of r.

Equation 5 gives fully efficient estimates of k and p. The equation may be solved by elementary methods (Haldane 1941) but the procedure is tedious and not ordinarily necessary. The procedure is to guess a value for k', evaluate both sides, and, by successive approximation and interpolation, carry the value of k' to the required number of decimal places, usually more than four. It is convenient to make initial approximation of k' by use of equation where P_0 is estimated as the ratio of obser zero observations to total observations. I further details, as well as a simple numeri example, the reader is referred to Haldane (194

A less tedious but not fully efficient proced is to estimate k by successive approximation satisfy equation 2. The method and its e ciency as compared to the maximum-likeliho method are discussed by Anscombe (1949, 1950

The negative-binomial distributions in t paper were first fitted using the log-likeliho equation 5 to estimate k. Some tabulation err were later discovered in the data and all of t distributions were recalculated by estimating from equation 2, with the exception of the dist butions for haddock in depth zones II and I (table 5) and for ocean perch (table 9). T values of k were not materially changed by usi the latter method which is, for these data, abo 90-percent efficient. Since the first term of t expected series, using equation 2, is determine there is one less degree of freedom than in usi equation 5.

As pointed out in the text, the negative nomial approaches the Poisson as k increas Some examples of the shape which the negati binomial may take as k increases with the mean fixed are shown in figure 7. A value of k about 0.1 would produce a curve similar in sha to the types found in this paper.

APPENDIX C

The logarithmic series

Fisher (1943) shows that the logarithmic ser arises from the negative binomial distributi when, in equation 2, appendix A, we let $k=$ write x for $p/(p+1)$, and replace the consta factor $(k-1)!$ in the denominator by a n constant factor, α, in the numerator. The pression for the expected number of species wi r individuals is then

$$\alpha x^r/r \quad \ldots\ldots\ldots\ldots\ldots\ldots$$

where r cannot now be zero. The successi terms of equation 2 are, of course, the requi series

$$\alpha x,\ \alpha x^2/2,\ \alpha x^3/3,\ \ldots\ldots\ldots\ldots$$

total number of species expected is

$$\sum_{r=1}^{\infty} \frac{\alpha}{r} x^r = -\alpha \log_e (1-x) \quad \ldots\ldots\ldots (3)$$

which it is seen that the distribution is ed to the algebraic expansion of the loga- n, hence its name.

he total number of individuals expected is

$$\sum_{r=1}^{\infty} \alpha x^r = \alpha x/(1-x) \quad \ldots\ldots\ldots\ldots (4)$$

ing S represent the total number of species rved and N the total number of individuals, nave from equations 3 and 4

$$S = -\alpha \log_e (1-x) \quad \ldots\ldots\ldots (5)$$

$$N = \alpha x/(1-x) \quad \ldots\ldots\ldots\ldots (6)$$

hat both α and x may be determined if S and re known. Fisher (1943) and Williams (1947) tables of $\log_{10} N/\alpha$ in terms of $\log_{10} N/S$ for ing equations 5 and 6, given S and N. The putations are straightforward and the param- s x and α, as well as the terms of the expected es (equation 2), may be readily determined a given set of data.

isher (1943) gives the following formula for variance of

$$V_\alpha = \frac{\alpha^3 \left\{ (N+\alpha)^2 \log_e \frac{2N+\alpha}{N+\alpha} - \alpha N \right\}}{(SN+S\alpha-N\alpha)^2} \quad \ldots\ldots (7)$$

formula enables one to calculate the standard r of α but Anscombe (1950) points out that ation 7 is appropriate only under certain ditions of sampling, these conditions not g completely met in the census sampling. combe gives the formula

$$V_\alpha \sim \frac{\alpha}{\log_e p} \quad \ldots\ldots\ldots\ldots (8)$$

ch appears to apply to the conditions met in census sampling, that is, comparisons of rvations on different sorts of biological ciations and comparisons of α from observa- s obtained in different years and with probable tive differences in abundance between areas years. The variances of α use in this paper estimated from equation 8.

APPENDIX D

A negative binomial distribution fitted to numbers of plaice eggs caught by a plankton net

The published data on sampling with plankton nets do not often give the total catches in successive tows, or sufficient data on number of tows with zero observations, to study the nature of the distributions encountered. Table 17 is a summary of the catches of plaice eggs in all stages of development in the southern part of the North Sea (Buchanan-Wollaston, 1923, table 3). The data were collected in 1914. Table 3 (Buchanan-Wollaston) indicates that 27 of the 50 stations yielded no plaice eggs in the plankton tows. The mean catch is 15.1 eggs with a standard deviation of 30.3.

Estimating k from formula 2, appendix B, we have

$$27/50 = (1+15.1/k)^{-k}$$

from which $k=0.1292$.

The expected variance from equation 1, appendix B, is

$$15.1 = 15.1^2/0.1292$$

or 1779.9, as compared to the observed variance, 915.5.

The observed frequency of plaice eggs per tow and the frequency expected on the hypothesis of a negative binomial distribution are presented in table 18. The probability of obtaining a worse fit by chance for $\Sigma\chi^2$ with three degrees of freedom is somewhat greater than 0.60, so that the hypothesis of a negative binomial distribution of the data may not be rejected

APPENDIX E

Proof, smaller sampling unit more efficient with heterogeneity present

When the variance is a function of the mean of a distribution, the effect of the size of sampling unit on the efficiency of sampling and on the precision of estimates cannot be expected to be obvious to the biologist. Experimental data confirming the validity of the efficiency of the smaller sampling unit, under conditions of heterogeneity, have been published by Fleming and Baker (1936), Beall (1939), and Finney (1946).

The precision of an estimate is defined as the reciprocal of its variance. Efficiency is defined in terms of the relative amounts of sampling required

to give estimates of equal precision. If, for example, we sample area A with sampling unit a_1, and then sample area $A/2$ with sampling unit a_2 obtaining an estimate equally precise, we say that sampling unit a_1 is one-half as efficient as sampling unit a_2.

Heterogeneity is defined as a nonrandom condition where the presence of an item at any given point tends to increase the probability of another item being in its immediate neighborhood. Heterogeneity is to be distinguished from another type of nonrandomness in which items, through competition or whatever cause, act to repel each other so that they tend to fill the available space equidistant from one another. With heterogeneity, as here defined, the variance is always greater than the mean.

The problem of sampling with different sizes of sampling units has been examined by Beall (1939) whose statement of the problem is followed closely in equations 1 through 7 below. It is hoped that the proofs which follow in equations 8 through 10 will be helpful in clarifying some of the confusion which appears to exist concerning the problem.

Consider N strata of equal size and a sampling unit of area k. If the total area of the ith stratum is A, the total number of sampling units it contains is $A/k=M$. M will, of course, vary with the size of the sampling unit.

Let X be the total number of fish in the N strata. Let u_{ij} be the number of fish in the jth sampling unit ($j=1,2, \ldots M$) in the ith stratum ($i=1,2, \ldots N$). Let u be the average number of fish per sampling unit calculated over the entire N strata. Then

$$X=MNu \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Within the ith stratum, let the mean value of u_{ij} be u_i and the variance

$$\sigma_i^2=\frac{1}{M-1}\sum_{j=1}^{M}(u_{ij}-u_i)^2 \quad \ldots\ldots\ldots (2)$$

Let m_i denote any number of samples taken in the ith stratum. Let the number of fish observed per sampling unit be $x_{i1}, x_{i2} \ldots x_{im_i}$ with mean x_i and estimated variance

$$S_i^2=\frac{1}{m_i-1}\sum_{j=1}^{mi}(x_{ij}-x_i)^2 \quad \ldots\ldots\ldots (3)$$

M being equal for all strata, Beall (1939) sta "The best linear estimate of X, that is, the mate with the minimum S. D. will be

$$F=M\sum_{i=1}^{N}x_i.''$$

The standard deviation will be, when m_i samp units have been drawn at random

$$\sigma_F=\sqrt{\left\{\sum_{i=1}^{N}\left(\frac{M_i(M_i-m_i)}{m_i}\sigma_i^2\right)\right\}} \ldots\ldots$$

where M_i is the total number of sampling u contained in the ith stratum.

Since the strata are all equal in area, $M_i=M$ $m_i=m$, so that equation 4 may be written

$$\sigma_F=\sqrt{\frac{M(M-m)}{m}\sum_{i=1}^{N}\sigma_i^2} \ldots\ldots\ldots$$

MN is the total number of sampling units. $MN=M_o$. mN is the total number of sam taken. Let $mN=m_o$. Substituting in equa 5 we have

$$\sigma_F=\sqrt{\frac{(M_o-m_o)M_o}{m_o}\cdot\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2} \ldots\ldots\ldots$$

Since M_o and m_o are fixed, the accuracy of estimate of F is determined by the average ance within strata:

$$\sum_{i=1}^{N}\frac{\sigma_i^2}{N}=\sigma_o^2$$

The sampling unit which gives the smallest v to σ_o^2 gives the greatest accuracy to F, the mate of X.

With the number of strata and the fractio the area to be examined fixed, σ_F may be supp to be affected by an increase in the size of sampling unit.

Let $\sigma_F=\sigma_F'$ with sampling unit $k=1$. $\sigma_F=\sigma_F''$ with sampling unit $k>1$. When k $M=M'$, and $m=m'$. When $k>1$, $M=m''$, $m=M''$. For each value of k, there will l value of σ_o^2 and $M''/M=m''/m=1/k$. By mentary algebra, we obtain from equation 5

$$\frac{\sigma_F'}{\sigma_F''}=\sqrt{\sigma_o'^2}\div\sqrt{\sigma_o''^2/k} \ldots\ldots\ldots$$

Proof that, with a Poisson distribution, all sizes sampling units are equally efficient.—Consider ith stratum. Let it contain M sampling units. the number of sampling units to be taken be Let F be the estimate of the total fish in the stratum. By (5):

$$\sigma_F = \sqrt{\frac{M(M-m)}{m}} \cdot \sigma_i^2 \quad \ldots\ldots\ldots (8)$$

nsider sampling unit $k' > k=1$. Since the a of the ith stratum is fixed, the total number sampling units of size k' is M/k'. For equal iency, it is required $\sigma_F^2 = \sigma_F'^2$ without increasing total area of the sample taken, so the number samples to be taken for $k' > k$ is m/k'. By ation 5:

$$\sigma_F{}^1 = \sqrt{\frac{M/k'(M/k' - m/k')}{m/k'}} \sigma_i'^2 \quad \ldots\ldots (9)$$

ere $\sigma_i'^2$ is the variance observed with the new pling unit, k'.

The comparative efficiency of the new sampling t will be equation (8) divided by equation (9):

$$\sigma_F^2/\sigma_F'^2 = \sigma_i^2 \div \sigma_i'^2/k' \quad \ldots\ldots\ldots (10)$$

The means observed may be considered pro- tional to the size of the sampling units used, vided sufficient samples are taken, since both sampling the same population density. So x'/k', where x' is the mean observed with pling unit $k' > k$. By definition of the distribution, $\bar{x} = \sigma_i^2$ and $x' = \sigma_i'^2 = k'x$. Substituting in equation 10:

$$\sigma_F^2/\sigma_F'^2 = \bar{x} \div k'\bar{x}/k' = 1.0$$

Proof that the smaller sampling unit is more efficient if heterogeneity is present.—With heterogeneity as defined above, the variance is greater than the mean. As with the Poisson distribution, the means observed will be proportional to the size of the sampling units used, so that $\bar{x} = \bar{x}'/k'$. Because of heterogeneity, however, the variances are no longer proportional to the size of the sampling unit as in the Poisson distribution where $\sigma_i^2 = \sigma_i'^2/k'$. For the negative binomial distribution and some other types of heterogeneous distributions, the relation between the mean and variance is of the type $\sigma^2 = a\bar{x} + b\bar{x}^2$. We have, therefore, $\sigma_i^2 < \sigma_i'^2/k'$ and it is immediately obvious from inspection of (10) that $\sigma_F^2/\sigma_F'^2$ will be less than 1.0.

It hardly needs to be pointed out that the application of this result to problems of heterogeneous sampling does not mean one's difficulties will be solved merely by taking the smallest possible sampling unit. Many practical factors will set a lower limit to the size of the sample it is desirable to take. In sampling with an otter trawl, for example, one would approach a point where more time would be consumed in lowering and raising the trawl than in fishing along the bottom. It must also be remembered that with the smaller sampling unit, minor sources of error may remain of the same magnitude and so become of proportionally greater importance with the smaller sampling unit.

REFERENCES CITED

ANSCOMBE, F. J.
1948. The transformation of Poisson, binomial and negative binomial data. Biometrika, vol. 35, p. 246.
1949. The statistical analysis of insect counts based on the negative binomial distribution. Biometrics, vol. 5, No. 2, pp. 165–173.
1950. Sampling theory of the negative binomial and logarithmic series distributions. Biometrika, vol. 37, p. 358.

BARTLETT, M. S.
1947. The use of transformations. Biometrics, vol. 3, No. 1, pp. 39–52.

BARNES, H.
1949a. A statistical study of the variation in vertical plankton hauls with special reference to the loss of the catch with divided hauls. Jour. Mar. Biol. Assn., N. S., vol. 28, p. 429.
1949b. On the volume measurement of water filtered by a plankton pump, with some observations on the distribution of planktonic animals. Jour. Mar. Biol. Assn., N. S., vol. 28, p. 651.
1951. On the variability of replicate plankton samples and some applications of "contagious" series to the statistical distribution of catches over restricted periods. Jour. Mar. Biol. Assn., N. S., vol. 30, p. 233.

BARNES, H., and T. B. BAGENAL.
1951. A statistical study of variability in catch obtained by short repeated trawls taken over an inshore ground. Jour. Mar. Biol. Assn., N. S., vol. 29, p. 649.

BEALL, GEOFFREY
1939. Methods of estimating the population of insects in a field. Biometrika, vol. 30, pp. 422–439.

BUCHANAN-WOLLASTON, H. J.
1923. The spawning of plaice in the southern part of the North Sea in 1913–14. Great Britain Min. Agric. and Fish., Fish. Invest., ser. II, vol. 5, No. 2, 36 pp.

COLE, LAMONT C.
1946. A theory for analyzing contagiously distributed populations. Ecology, vol. 27, No. 4, pp. 329–341.

CURTIS, J. H.
1943. On transformations used in analysis of variance. Ann. Math. Stat., vol. 14, p. 10.

FINNEY, D. J.
1946. Field sampling for the estimation of wireworm populations. Biometrics, vol. 2, No. 1, pp. 1–7.

FISHER, R. A.
1941. The negative binomial distribution. Ann. Eug., vol. 11, pp. 182–187.

FISHER, R. A., A. S. CORBET, and C. B. WILLIAMS.
1943. The relation between the number of individuals and the number of species in a random sample of an animal population. Jour. Animal Ecol., vol. 12, No. 1, pp. 42–58.

FLEMING, W. E., and F. E. BAKER
1936. A method for estimating populations of larvae of the Japanese beetle in the field. Jour. Agric. Res., vol. 53, No. 5, pp. 319–341.

GREENWOOD, M., and G. U. YULE
1920. An inquiry into the nature of frequency distr
tions representative of multiple happenings with
ticular reference to the occurrence of multiple att
of disease or of repeated accidents. Jour. Roy. S
Soc., vol. 83, p. 255.

HALDANE, J. B. S.
1941. The fitting of binomial distributions. Ann. E
vol. 11, p. 179.

HICKLING, C. F.
1933. The natural history of the hake. Part IV.
determination and growth-rate. Great Britain
Agric. and Fish., Fish. Invest., ser. II, vol. 13, N
120 pp.

KENDALL, D. G.
1948. On some modes of population growth leadin
R. A. Fisher's logarithmic series distribution.
metrika, vol. 35, p. 6.

NEYMAN, J.
1939. On a new class of contagious distributions,
plicable in entomology and bacteriology.
Math. Stat., vol. 10, p. 35.

POLYA, G.
1931. Sur quelques points de la theorie des probabili
Ann. Inst. Poincaré, vol. 1, p. 117.

QUENOUILLE, M. H.
1949. A relation between the logarithmic, poisson,
negative binomial series. Biometrics, vol. 5, N
pp. 162–164.

SILLIMAN, R. P.
1946. A study of variability in plankton townet cat
of Pacific pilchard, *Sardinops caerulea*, eggs.
Mar. Res. vol. 6, pp. 74–83.

THOMAS, M.
1949. A generalization of Poisson's binomial limit
use in ecology. Biometrika, vol. 36, p. 18.

THOMPSON, H.
1928. Haddock biology. V. General features in
biology of the haddock (*Gadus aeglefinus* L.) in
landic waters in the period 1903–1926. Fishery
Scotland, Sci. Invest. No. 5, 63 pp.

WILLIAMS, C. B.
1944. Some applications of the logarithmic series
the index of diversity to ecological problems.
Ecology, vol. 32, No. 1, pp. 1–44.
1947. The logarithmic series and its applicatio
biological problems, Jour. Ecology, vol. 34, N
pp. 253–272.

WINSOR, C. P. and G. L. CLARKE
1940. A statistical study of variation in the cat
plankton nets. Jour. Mar. Res., vol. 3, No. 1
1–34.

WINSOR, C. P., and L. A. WALFORD.
1936. Sampling variations in the use of plankton
Jour. Cons. Int. Expl. Mer., vol. 11, pp. 190–20

TABLES

BLE 1.—*Distribution of values of χ^2 for paired tow tches of haddock, whiting, common skate, and total fish*

[Only hauls with $n_1+n_2>9$ are included]

χ^2	Proportion of cases expected	Haddock	Whiting	Common skate	Total fish
)16	0.1	1	1	1	3
)64	0.1		1	2	1
148	0.1	2	1	4	3
275	0.1	1	1	1	1
455	0.1	2		3	
708	0.1			1	4
)74	0.1	1	3		3
342	0.1	2	2		4
706	0.1		5	1	2
335	0.09	8	4	4	6
r 6.635	0.01	33	42	13	82
Total cases		50	60	30	109

BLE 2.—*Values of total and mean χ^2 for haddock tows of different sizes*

[Only hauls with $n_1+n_2>9$ are included]

Number of fish	Number of paired tows	$\Sigma\chi^2$	Mean χ^2
19	19	184.4	9.70
99	8	120.3	15.04
-149	3	181.7	60.57
-199	4	187.4	46.95
-299	2	25.7	12.85
-399	2	90.1	45.05
-499	3	452.7	150.90
-599			
-699	2	142.6	71.30
-799	2	116.0	108.00
3	1	168.9	168.90
3	1	1.5	1.50
4	1	49.1	49.10
4	1	176.8	176.80
1	1	885.2	885.20

BLE 3.—*Distribution of catches per tow of haddock at two-tow stations*

Number per tow	Number of tows		Total
	Tow 1	Tow 2	
	25	26	51
	8	5	13
	4	3	7
	3	5	8
	1	2	3
	4	1	5
	1	2	3
	1	3	4
	1	1	2
		1	1
	1	2	3
	1	1	2
	1	1	2
	2	1	3
		1	1
		1	1
	1	1	2
	1		1
5	2	2	4
0	3	1	4
5	2	2	4
0	3	2	5
5		2	2
r 45	23	22	45
Total	88	88	176

TABLE 4.—*Observed catch of haddock at two-tow stations and catch expected on the hypothesis of a negative binomial distribution*

Number of haddock/tow	Number of tows observed	Number of tows expected	χ^2
0	51	50.98	0.000
1	13	15.14	.302
2–3	15	10.39	2.045
4–7	15	11.21	1.281
8–14	13	11.15	.307
15–25	9	10.64	.253
43–4,586	47	66.49	5.714
Total	176	176.00	10.592

TABLE 5.—*Observed catch of haddock and catch expected on the hypothesis of a negative binomial distribution, depth zones, I, II, and III*

Depth zone	Number of haddock/tow	Number of tows observed	Number of tows expected	χ^2
I	0	35	34.99	0.000
	1–4	14	11.42	.583
	5–20	14	10.81	.942
	21–70	7	10.09	.946
	71–1,174	14	16.69	.434
Total		84	84.00	2.905
II	0	37	41.50	0.488
	1–5	28	26.04	.148
	6–10	16	10.46	2.934
	11–18	11	10.54	.021
	19–30	15	10.64	1.787
	31–47	9	10.25	.152
	48–71	13	10.03	.879
	82–4,586	43	52.54	1.732
Total		172	172.00	8.141
III	0	33	33.01	0.000
	1	11	11.41	.015
	2–3	15	13.40	.191
	4–6	16	12.39	1.054
	7–10	5	10.90	3.194
	11–16	9	10.91	.334
	17–25	13	10.41	.644
	26–42	13	10.41	.644
	43–91	9	10.01	.102
	91–161	3	4.15	.319
Total		127	127.00	6.497

TABLE 6.—*Observed catch of common skate* (Raja erinacea) *and catch expected on the hypothesis of a negative binomial distribution, depth zones I, II, and III combined*

Number of skates/tow	Number of tows observed	Number of tows expected	χ^2
0	195	195.00	0.000
1–2	54	60.71	.742
3–4	26	27.93	.098
5–6	24	18.06	1.967
7–8	14	13.09	.136
9–10	12	10.06	.374
11–13	7	11.45	1.729
14–17	16	10.93	2.352
18–23	12	10.90	.111
24–33	8	10.43	.566
34–87	16	14.44	.169
Total	383	383.00	8.244

TABLE 7.—*Observed catch of whiting* (Merluccius bilinearis) *and catch expected on the hypothesis of a negative binomial distribution, depth zones II and III*

Number of whiting/tow	Number of tows observed	Number of tows expected	χ^2
0	35	35.00	0.000
1	5	12.62	4.601
2–3	12	15.43	.762
4–5	6	10.71	2.071
6–8	10	12.20	.397
9–12	8	12.53	1.638
13–16	12	10.12	.349
17–21	9	10.55	.228
22–27	14	10.63	1.068
28–34	16	10.49	2.894
35–42	8	10.25	.299
43–52	16	10.95	2.329
53–2,690	148	138.52	.649
Total	299	299.00	17.285

TABLE 8.—*Observed catch of ocean perch* (Sebastes marinus) *and catch expected on the hypothesis of a negative binomial distribution, depth zone III*

[Includes 61 tows made in subareas adjacent to Georges Bank proper.]

Number of ocean perch per tow	Number of tows observed	Number of tows expected	χ^2
0	59	60.24	0.026
1–2	15	13.84	.097
3–6	14	10.76	.976
7–14	8	10.29	.509
15–30	12	10.46	.227
31–60	14	10.24	1.380
61–890	26	32.17	1.183
Total	148	148.00	4.398

TABLE 9.—*Summary of tables 4–8, with degrees of freedom and probability associated with each* $\Sigma\chi^2$

Species	Depth zone	$\Sigma\chi^2$	d. f.[1]	Probability greater than
Haddock [2,3]	I–III	10.592	5	0.05
Haddock	I	2.905	2	.20
Haddock	II	8.141	6	.20
Haddock	III	6.497	8	.50
Common skate	I–III	8.244	8	.30
Whiting	II–III	17.285	10	.05
Ocean perch [3]	III	4.398	5	.40

[1] See appendix B.
[2] Two-tow station data.
[3] Includes some tows made in subareas adjacent to Georges Bank proper.

TABLE 10.—*Number of tows, mean, variance, and stan deviation, for haddock catches by subarea and depth zo*

Subarea	Depth zone	Number of tows	Mean	Variance	Sta de ti
G	I	10	55.6	3,094	
	II	8	19.3	255	
	III	42	27.6	1,391	
H	I	14	77.2	31,216	
	II	14	53.3	11,015	
	III	29	14.3	415	
J	I	4	32.3	2,190	
	II	24	230.3	123,956	
	III	17	18.2	181	
M	I	18	46.3	8,720	
	II	42	288.0	633,397	
	III	11	4.1	39	
N	I	15	87.2	59,701	
	II	39	68.5	13,717	
	III	18	0.1	0	
O	I	23	51.1	3,532	
	II	45	16.5	1,153	
	III	9	0.1	0	

TABLE 11.—*Number of tows, mean, variance, and stan deviation, for whiting catches by subarea and depth zo*

Subarea	Depth zone	Number of tows	Mean	Variance	Sta de ti
G	I	10	57.9	14,038	
	II	8	135.0	36,288	
	III	42	75.6	5,050	
H	I	14	235.5	108,481	
	II	14	206.6	46,361	
	III	29	153.9	35,267	
J	I	4	164.8	18,644	
	II	24	129.8	115,563	
	III	17	154.6	38,841	
M	I	18	60.8	3,012	
	II	42	158.7	25,358	
	III	11	376.3	256,151	
N	I	15	113.7	13,951	
	II	39	100.0	27,781	
	III	18	187.8	395,583	
O	I	23	134.9	42,828	
	II	45	78.6	7,761	
	III	9	13.4	294	

TABLE 12.—*Frequency of occurrence of species with diff numbers of individuals and expected frequency of oc rence on the hypothesis of a logarithmic series distribu 1948–50*

a. DEPTH ZONE I

Number of individuals	Species observed	Species expected	χ^2	d. f.	Pr abi gre tha
1–4	12	10.180	0.325		
5–12	2	4.981	1.784		
13–34	4	4.947	.181		
35–96	4	4.989	.196		
97–250	6	4.605	.423		
251–10,433	15	14.298	.034		
Total	44	44.000	2.943	4	

E 12.—*Frequency of occurrence of species with different nbers of individuals and expected frequency of occurce on the hypothesis of a logarithmic series distribution, 8–50*—Continued

EPTH ZONE II

ber of individuals	Species observed	Species expected	χ^2	d. f.	Probability greater than—
	10	10.453	0.020		
	7	5.113	.696		
	5	5.086	.001		
	4	5.032	.212		
	5	5.037	.000		
367	17	17.279	.004		
Total	48	48.000	0.933	4	0.90

EPTH ZONE III

ber of individuals	Species observed	Species expected	χ^2	d. f.	Probability greater than—
	13	10.175	0.784		
	6	4.905	.244		
	7	5.139	.674		
	4	5.033	.212		
	7	5.001	.799		
893	12	18.747	2.428		
tal	49	49.000	5.141	4	0.30

E 13.—*Frequency of occurrence of species with different nbers of individuals and expected frequency of occurce on the hypothesis of a logarithmic series distribution, th zones I–III*

EAR 1948

ber of individuals	Species observed	Species expected	χ^2	d. f.	Probability greater than—
	8	10.658	0.663		
	6	5.118	0.152		
	4	5.015	.205		
	5	5.004	.000		
667	19	16.205	.482		
Total	42	42.000	1.502	3	0.60

EAR 1949

ber of individuals	Species observed	Species expected	χ^2	d. f.	Probability greater than—
	15	10.319	2.124		
	4	4.977	0.192		
	4	4.948	.182		
	4	5.060	.222		
	3	4.997	.798		
755	22	21.699	.004		
Total	52	52.000	3.522	4	0.40

TABLE 13.—*Frequency of occurrence of species with differen numbers of individuals and expected frequency of occurrence on the hypothesis of a logarithmic series distribution, depth zones I–III*—Continued

c. YEAR 1950

Number of individuals	Species observed	Species expected	χ^2	d. f.	Probability greater than—
1–3	9	9.923	0.086		
4–8	4	4.786	.129		
9–21	4	5.013	.205		
22–54	5	5.021	.000		
55–18,496	28	25.257	.298		
Total	50	50.000	0.718	3	0.80

TABLE 14.—*Individuals (N), species (S), indices of diversity (α), and variance of α by depth zones and years*

Year	Depth zone	N	S	α	V_α
1948	I	5,837	27	3.661	0.4965
	II	16,496	34	4.096	.4934
	III	15,315	34	4.138	.5036
1949	I	11,246	33	4.178	.5290
	II	31,353	40	4.523	.5114
	III	15,138	37	4.564	.5630
1950	I	22,610	38	4.453	.5218
	II	17,610	35	4.195	.5029
	III	7,570	34	4.589	.6194
1948–50	I	39,693	44	4.888	
	II	65,459	48	5.018	
	III	38,024	49	5.551	

TABLE 15.—*Distribution of number of species caught per tow in each year (A) and the same with elimination of the first three tows following a change of net (B)*

Number of species per tow	Number of tows							
	1948		1949		1950		1951	
	A	B	A	B	A	B	A	B
0	1	1					3	
1	10	9						
2	12	7					3	1
3	11	10	5	5	2	1	1	1
4	11	9	5	1	6	5	9	6
5	10	5	9	5	5	5	8	6
6	18	17	13	11	10	8	7	5
7	17	15	13	10	11	10	5	5
8	12	11	11	7	15	14	10	10
9	5	4	18	17	14	13	15	15
10	13	11	9	5	12	9	8	8
11	11	10	14	10	17	13	12	11
12	3	3	7	5	10	8	5	5
13	4	3	11	8	9	9	6	6
14	2	1	4	3	1	1	3	3
15	2	1	2		1	1	1	
16			1	1	4	4	2	2
17					1	1	1	1
18					1	1		
Total	142	117	122	88	119	103	99	85
Mean	6.468	6.573	8.787	8.761	9.328	9.379	8.414	9.950

TABLE 16.—*Distribution of species and tows and the expected number of tows on a hypothesis of a Poisson distribution, 1948–51*

a. YEAR 1948

	Number of Species									
	0–3	4	5	6	7	8	9	10	11 and more	Total
Observed tows	27	9	5	17	15	11	4	11	18	117
Expected tows	12.51	12.72	16.72	18.32	17.20	14.13	10.32	6.78	8.30	117
x^2	16.780	1.088	8.215	0.095	0.281	0.693	3.870	2.627	11.340	44.989

b. YEAR 1949

	Number of Species									
	0–4	5	6	7	8	9	10	11	12	13 and more	Total
Observed tows	6	5	11	10	7	17	5	10	5	12	88
Expected tows	5.60	5.94	8.66	10.84	11.98	11.56	10.13	8.06	5.89	9.33	88
x^2	0.029	0.149	0.632	0.065	2.070	2.560	2.598	0.467	0.134	0.764	9.468

c. YEAR 1950

	Number of Species									
	0–5	6	7	8	9	10	11	12	13	14 and more	Total
Observed tows	11	8	10	14	13	9	13	8	9	8	103
Expected tows	9.74	8.23	11.02	12.92	13.47	12.63	10.77	8.41	6.07	9.74	103
x^2	0.163	0.006	0.094	0.090	0.018	1.043	0.462	0.026	1.414	0.311	3.627

d. YEAR 1951

	Number of Species									
	0–5	6	7	8	9	10	11	12	13 and more	Total
Observed tows	14	5	5	10	15	8	11	5	12	85
Expected tows	9.60	7.63	9.83	11.12	11.18	10.01	8.32	6.27	11.04	85
x^2	2.016	0.907	2.377	0.113	1.216	0.203	0.865	0.257	0.083	8.037

TABLE 17.—*Observed frequency of plaice eggs in 5(... with a plankton net*

[Data from Buchanan-Wollaston, 1923]

Number of plaice eggs	Number of tows observed	Number of plaice eggs	N... o... ob...
0	27	36	
1	3	38	
2	1	54	
4	3	56	
5	1	72	
8	2	95	
14	1	121	
18	1	123	
22	1		
33	1	Total	
34	1		

TABLE 18.—*Observed and expected frequencies of plai... in 50 tows with a plankton net*

Number of plaice eggs	Number of tows observed	Number of tows expected
0	27	27.00
1	3	3.46
2–5	5	5.23
6–16	3	4.94
17–50	6	4.94
Over 50	6	4.43
Total	50	50.00